

University of Wollongong

## Research Online

---

Faculty of Engineering and Information  
Sciences - Papers: Part A

Faculty of Engineering and Information  
Sciences

---

1-1-2013

### Bayesian nonparametric reliability analysis for a railway system at component level

Payam Mokhtarian

*University of Wollongong*, [payam@uow.edu.au](mailto:payam@uow.edu.au)

Mohammad-Reza Namazi-Rad

*University of Wollongong*, [mrاد@uow.edu.au](mailto:mrاد@uow.edu.au)

Tin Kin Ho

*University of Wollongong*, [markho@uow.edu.au](mailto:markho@uow.edu.au)

Thomas Suesse

*University of Wollongong*, [tsuesse@uow.edu.au](mailto:tsuesse@uow.edu.au)

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

---

#### Recommended Citation

Mokhtarian, Payam; Namazi-Rad, Mohammad-Reza; Ho, Tin Kin; and Suesse, Thomas, "Bayesian nonparametric reliability analysis for a railway system at component level" (2013). *Faculty of Engineering and Information Sciences - Papers: Part A*. 1906.  
<https://ro.uow.edu.au/eispapers/1906>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Bayesian nonparametric reliability analysis for a railway system at component level

## Abstract

Railway system is a typical large-scale complex system with interconnected sub-systems which contain numerous components. System reliability is retained through appropriate maintenance measures and cost-effective asset management requires accurate estimation of reliability at the lowest level. However, real-life reliability data at component level of a railway system is not always available in practice, let alone complete. The component lifetime distributions from the manufacturers are often obscured and complicated by the actual usage and working environments. Reliability analysis thus calls for a suitable methodology to estimate a component lifetime under the conditions of a lack of failure data and unknown and/or mixture lifetime distributions. This paper proposes a nonparametric Bayesian approach with a Dirichlet Process Mixture Model (DPMM) to facilitate reliability analysis in a railway system. Simulation results will be given to illustrate the effectiveness of the proposed approach in lifetime estimation.

## Disciplines

Engineering | Science and Technology Studies

## Publication Details

Mokhtarian, P., Namzi-Rad, M., Ho, T. Kin. & Suesse, T. (2013). Bayesian nonparametric reliability analysis for a railway system at component level. IEEE International Conference on Intelligent Rail Transportation (ICIRT) (pp. 197-202). China: The Institute of Electrical and Electronics Engineers Inc.

# Bayesian Nonparametric Reliability Analysis for a Railway System at Component Level

Payam Mokhtarian<sup>†\*</sup>, Mohammad-Reza Namazi-Rad<sup>\*</sup>, Tin Kin Ho<sup>\*</sup>, and Thomas Suesse<sup>†</sup>

<sup>†</sup>National Institute for Applied Statistics Research Australia, University of Wollongong, NSW 2522, AUSTRALIA

<sup>\*</sup>Smart Infrastructure Facility, University of Wollongong, NSW 2522, AUSTRALIA

**Abstract**—Railway system is a typical large-scale complex system with interconnected sub-systems which contain numerous components. System reliability is retained through appropriate maintenance measures and cost-effective asset management requires accurate estimation of reliability at the lowest level. However, real-life reliability data at component level of a railway system is not always available in practice, let alone complete. The component lifetime distributions from the manufacturers are often obscured and complicated by the actual usage and working environments. Reliability analysis thus calls for a suitable methodology to estimate a component lifetime under the conditions of a lack of failure data and unknown and/or mixture lifetime distributions. This paper proposes a nonparametric Bayesian approach with a Dirichlet Process Mixture Model (DPMM) to facilitate reliability analysis in a railway system. Simulation results will be given to illustrate the effectiveness of the proposed approach in lifetime estimation.

**Keywords**—Finite Mixture Model; Lifetime Estimation; Nonparametric Bayesian; Reliability Modeling

## I. INTRODUCTION

Rail system requires high asset investment and yields low return over the long asset life cycle. It is a complex system with physically interconnected and functionally interdependent sub-systems and components, such as tracks, rolling stocks, power supply and signaling. The overall reliability is imperative to the quality of service provision and it is upheld through appropriate maintenance works. Maintenance scheduling is a delicate balancing act between cost and reliability. The desired level of reliability is the driver while the cost is the constraint. The system reliability inevitably relates to that of the sub-systems and components through the system configuration and function criticality.

In order to evaluate system reliability, it is essential to understand the reliability at the lowest levels. However, not every sub-system or component comes with adequate reliability data when its condition changes, usually deteriorates, due to usage, tear-and-wear, fatigue and working conditions. Failure data is not particularly well recorded, and in most cases, it is simply not available as rail systems tend to be over-maintained to eliminate failures at all. Failure behavior of the components is not necessarily constant or homogeneous. It may change over time because of possible maintenance regimes, service intensity, operation conditions, locations and climate, and vary over different components. These factors attribute to an unknown component lifetime distribution or a mixture of distributions, which complicates the estimation of component lifetime and thus fails to inform the necessary maintenance planning. To address the uncertainties on

component lifetime estimation, nonparametric statistical approaches are conceived to be a useful tool to extract lifetime information from limited available data [1].

Reliability analysis is always related to statistical approaches as the commonly adopted lifetime models are usually expressed in probability density functions [1]. Applications in railway systems have not been very extensive but successful examples can be found from component to system levels [2-4]. In order to estimate the component lifetime at a particular time period with limited real-life data and uncertain lifetime distribution, a nonparametric Bayesian approach at sub-system or component level is proposed here. Bayesian models have been employed in various railway system reliability studies [5-7], particularly in response to the uncertainty in the condition deterioration of the system or component through its life-cycle.

With Bayesian models, statistical inference can be built up from little knowledge on the component failure data and distributions, and it evolves by incorporating additional data whenever it is made available. Bayesian methods are broadly classified into parametric and nonparametric approaches. The former has the advantage of simple representation, in the sense that model parameters are able to explain the behavior of the entire data. However, the resulting model strongly depends on stringent model assumptions and imposes certain structural restrictions. The latter is quite commonly adopted in practice when the model assumptions do not always hold or the available data does not contain sufficient information.

As the component lifetime distribution in railway may be a composition of a number of unknown distributions, a mixture distribution, instead of a typical one such as Weibull and Lognormal, is a more realistic model. A Bayesian nonparametric method, based on Dirichlet Process Mixture Model (DPMM) using Markov Chain Monte Carlo (MCMC) algorithm, is proposed here. DPMM allows an empirical mixture distribution to fit the available failure data. The number and characteristics of the mixtures may be unknown but they can be captured through gradual feeding of available data [8-11]. In addition, different kernel distributions of the model are possible and the comparison of the estimation capability will be discussed through simulation. The main objective of this study is to find out the effectiveness of nonparametric Bayesian methods in the estimation of the component reliability and the necessary conditions of the available data to achieve such effectiveness.

The remainder of this paper is structured as follows. In Section II, the nonparametric Bayesian methods and Dirichlet

process applied in reliability analysis is reviewed. Section III describes the methodology and MCMC algorithm for estimating the lifetime mixture density function based on the Bayesian nonparametric method and DPMM. In Section IV, data based on the nature of the realistic lifetime data is generated to examine performance of the proposed method, followed by result analysis. Finally, Section V concludes the paper with a summary of our major findings and a discussion of avenues for future researches.

## II. NONPARAMETRIC BAYESIAN METHOD IN RELIABILITY ANALYSIS

The nature of railway lifetime data is similar to lifetime data in other domains as the distribution of survival time and failure rates is similar. The general concept of survival analysis can be applied to the railway reliability analysis. Reliability of a component in a system is defined as the probability that it can perform adequately under the expected operation condition over a specific period of time [1]. This probability can be expressed as

$$R(t) = P(T > t) = \int_t^{\infty} f(x)dx \quad (1)$$

where  $f(x)$  is the Probability Density Function (PDF) (or time-of-failure density function) of failure time,  $T$  is a random variable denoting the failure time, and  $R(t)$  is the reliability function. The failure rate (or hazard rate) is defined as the rate of possible failures for the survivors to time  $t$ . It is denoted as

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)} \quad (2)$$

where  $F(t)$  is the Cumulative Distribution Function (CDF), which is the probability that the component may fail within time  $t$ , i.e.

$$F(t) = 1 - R(t) = P(T \leq t) \quad (3)$$

In reliability analysis, inference is based on  $f$ , therefore estimates of the parameters governing  $f$  or an accurate approximation of  $f$  are required. To estimate these unknown quantities, various statistical methods have been proposed [12].

In practice, it is not necessary to consider a homogenous behavior of a failure time distribution. Due to usage and condition, the component lifetime in a railway system is likely best captured by a multi-component lifetime distribution. Such distribution is often called mixture distribution which is a weighted sum of distributions, weighted by probability sum of one [13]. A mixture distribution can be written as

$$f(t) \sim \sum_{k=1}^K \pi_k f_k(t|\theta_k) \quad (4)$$

where  $f(t)$  can be considered as a composition of a finite number of distributions  $f_k$ , each specified by parameters  $\theta_k$ .  $f_k$ 's are commonly called mixture components and often the same component distribution with different parameters is used, i.e.  $f_1 = f_2 = \dots = f_K$ , but  $\theta_k$ 's are different. The parameter  $\pi_k$  is the  $k$ -th mixture coefficient or probability of the influence weight of the  $k$ -th mixture component. For this finite mixture model, the number of components  $K$  has to be fixed or estimated. Equation (4) represents a parametric mixture model.

For reliability analysis, both parametric and nonparametric mixture density estimations have been applied [14].

The parametric methods rely on the restrictive parametric assumptions of the failure time distribution. In certain situations, it may be difficult to specify a parametric model for the failure time distribution [1][12]. Reliability analysis in railway systems usually requires a large number of failure data, which may not be available for all components due to time and cost limitations. Furthermore, the failure mechanisms of certain sub-systems may be unknown and they involve multiple components or steps which are difficult to model by a simple lifetime distribution. Therefore, it is not always possible to assign a parametric distribution for the failure mechanisms in complex systems or mixture components, especially during the design and development stages. Thus, a more flexible and generic method needs to be employed.

Nonparametric Bayesian methods can offer significant modeling flexibility because no restrictive parametric assumptions of the lifetime distribution are required and the analysis can be conducted with smaller sample size compared to the non-Bayesian approach [15]. With the development of computation techniques, such as Markov Chain Monte Carlo (MCMC) methods, the Bayesian nonparametric data analysis became more practical. Since the lifetime distribution of a complex system often involves a number of failure distributions, a mixture model is useful and necessary to lead to a more accurate estimation [16]. Moreover, the number of components  $K$  of a mixture failure distribution is unknown in practice. The main focus in this paper is on the Dirichlet Process Mixture Model (DPMM) with Bayesian nonparametric perspective that also adjusts automatically for the unknown  $K$ , whereas most of the parametric and other nonparametric approaches do not [16]. Indeed, DPMM is robust on capturing the components of a mixture failure distribution.

### A. Nonparametric Bayesian Approach

Bayesian nonparametric models have been widely applied in system reliability measurement since its emergence. In any Bayesian analysis with non-informative priors, the likelihood function dominates the priors [17], which is similar to the non-Bayesian approach where estimation is fully based on likelihood. This is a motivation for using Bayesian nonparametric methods for the reliability measurements. Unlike parametric methods with fixed parametric forms, Bayesian nonparametric methods are well suited for unknown reliability functions. The flexibility of having infinite parameters is presented by avoiding critical dependency on prior parameters [18], the Bayesian nonparametric approach hence offers significant applicability on model selection especially for data of mixed types. The procedure of nonparametric data analysis is often considered as a distribution-free method [15] and the distribution of the data may be based on the rank of which the distribution is most influenced by the data. However, when both parametric and nonparametric methods are applicable to a problem, the parametric method is generally preferred due to its efficiency and simplicity. When the assumptions for the parametric method are questionable, nonparametric methods are more applicable [15].

Recently, the successful applications of nonparametric Bayesian methods in reliability analysis have increased remarkably. The Bayesian approach has been applied to the estimation of the derivatives of the cumulative hazards for a multiplicative counting process model [19]. Further, the survival analysis employs Bayesian nonparametric models to evaluate the proportional hazard function [20]. Moreover, an adapted version of MCMC algorithm has been used to simulate Bayesian nonparametric models for hazard functions [12][21].

### B. Dirichlet Process Mixture Model (DPMM)

Dirichlet Process (DP) mixture models have been applied in a wide range of applications in the area of Bayesian nonparametric data analysis [22]. A DP mixture model can handle random mixing distributions and is not restricted to one specific parametric distribution. One fundamental motivation for using DP construction is that the posterior distribution can be obtained easily [16]. Moreover, a number of studies in Bayesian reliability analysis have been suggested to use DP as prior [12, 16, 23].

In a Bayesian framework, it is necessary to specify a prior distribution to obtain, via Bayes' theorem, the posterior distribution on which statistical inference on the data is based on. The Dirichlet Process prior has been widely applied in Bayesian nonparametric approaches since its initial application [20]. The DP prior fits rich classes of Bayesian nonparametric models with its fulfillment of the two properties proposed in [22]. First, it is flexible in support of prior distributions and the posteriors can be tractably analyzed. Second, it can capture the number  $K$  of unknown mixture components. Moreover, the "Gibbs Sampler" algorithm [24] provides a computationally attractive tool overcoming otherwise difficulties associated with sampling from the posterior distribution. This leads to an increased popularity of DP mixture models in Bayesian nonparametric data analysis.

The Dirichlet distribution is the first step when implementing the Dirichlet Process Mixture Model (DPMM) in nonparametric Bayesian analysis. It is often used as the prior distribution in Bayesian inference because it is the conjugate prior of the multinomial distribution. Its probability density function is defined as the probabilities of  $n$  discrete variables  $X = \{X_1, X_2, \dots, X_n\}$  are  $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , i.e.,  $P(X = X_i) = \theta_i$ , while  $\theta$  follows a Dirichlet distribution  $\theta \sim \text{Dir}(\theta; v, M)$ , in which the PDF is defined as

$$P(\theta|v, M) = Q(v) \prod_{i=1}^n \theta_i^{vm_i-1} \quad (5)$$

where  $M = \{m_1, m_2, \dots, m_n\}$  is the base measure of  $X$ , and  $v$  is a parameter showing how concentrated the probability would be around  $M$ . When the value of  $v$  is 1, the Dirichlet distribution is closer to a uniform distribution. When  $v$  is below 1, it will show that most probabilities will be concentrated in a few ones [16]. Here,  $Q(v)$  is the normalizing constant expressed in terms of the gamma function as follows

$$Q(v) = \frac{\Gamma(v)}{\prod_{i=1}^n \Gamma(vm_i)} \quad (6)$$

The second step is to introduce the Dirichlet Process (DP) based on the Dirichlet distribution. The DP can be considered

as flexible continuous case of the Dirichlet distribution. The definition of the DP includes two base parameters. The first one is a positive scalar (concentration) parameter  $v$ , which expresses the belief towards  $G$  and the second one is a probability base distribution  $G_0$ , which is a nonparametric distribution (empirical distribution). The  $G$  is said to follow a Dirichlet Process, which can be written as

$$G \sim \text{DP}(vG_0) \quad (7)$$

$G$  is a discrete distribution containing values drawn from  $G_0$ , while  $G_0$  is a base distribution instead of a base measure  $M$ . It should be noted that DP is a "distribution over distributions" [16].

In order to implement the DPMM, we apply a mixture model, as in (4), on the DP. For general DP mixture models, when the data size grows and the data becomes more complicated, the theory dictates to assign an infinite number  $K$  of mixture components and parameters  $\theta$  growing with the data, so

$$t_i \sim f_i(\theta_i), i = 1, 2, 3, \dots \quad (8)$$

If  $G$  is set as a DP prior, it becomes the Dirichlet Process Mixture Model. Thus, the complete hierarchical form of the DPMM is:

$$\begin{aligned} t_i &\sim f_i(\theta_i) \\ \theta_i &\sim G(\cdot) \\ G &\sim \text{DP}(vG_0) \end{aligned} \quad (9)$$

Here,  $G_0$  is an infinite-dimensional distributional parameter, which makes the DPMM a nonparametric method. The behavior of the DP mixture model, however, is sensitive to the choice of the parameters [23], e.g.  $G_0$ , which is affected by the choice of the kernel distribution  $f_i$ . The basic choice is the DP Gaussian mixture model (DPGMM) with Gaussian distribution as the kernel distribution because of its flexible applications in both conjugate and non-conjugate distributions [25]. However, as the property of lifetime data requires the failure time to be positive ( $t > 0$ ), DPGMM with the data range over all real numbers is not suitable for reliability analysis. In this paper, other asymmetric kernel distributions, such as the Exponential, Weibull and Lognormal distribution, are considered for their flexibility and efficiency to fit mixture models. In particular, the results upon using the Lognormal kernel as the base distribution will be presented in Section V where comparison results are provided.

### C. Choice of DP parameters for DPMM

The Dirichlet Process is a very flexible and powerful due to two parameters  $v$  and  $G_0$ . The form of the base measure  $G_0$  plays an essential role towards model performance. It is hard to decide on the base measure because it is heavily dependent on the kernel distribution. The choice of  $G_0$  is conducted by mathematical convenience. Conjugate distributions are preferred for computational feasibility. Thus, the Dirichlet Process Gaussian Mixture Model (DPGMM) with both conjugate and non-conjugate base distributions has been widely applied [16, 24, 26]. A Bayesian nonparametric model using mixtures of Weibull distributions is developed in a previous study [11]. This is mixing on both the shape and scale parameters of the Weibull kernel. Several benefits for using a

Weibull kernel have been illustrated. One is that it allows hazards to increase more rapidly than other candidates, and another is that its survival function is computationally available in closed form. Furthermore, a MCMC algorithm to fit the model using both uncensored and right censored data is used.

The scalar parameter  $\nu$  controls how close the process is to the base distribution  $G_0$ , which crucially determines model performance. It was pointed out that the magnitude of  $\nu$  represents the degree of faith in the base distribution, which can be expressed in the formula as the number of clusters around the base distribution [16].

### III. METHODOLOGY AND ALGORITHM

The Dirichlet Process depends heavily on its base distribution  $G_0$ . Based on the nonparametric statistical properties, an empirical cumulative density function (CDF) of  $G_0$  is obtained as well as its probability density function (PDF) [20]. Consequently, the estimated CDF, PDF and hazard rate function (HRF) of lifetime data can be obtained by the hierarchical structure depicted in (9). In this section, a procedure for estimating the mixture density function of lifetime data using DPMM based on the Lognormal kernel function is given.

#### A. Model Specifications

To estimate the empirical distribution of  $G_0$ , a kernel estimating method is adopted, with Lognormal as the kernel distribution [20]. The Lognormal distribution has the CDF of

$$\Lambda_{LN}(t|\mu, \sigma^2) = \Phi\left[\frac{\log(t) - \mu}{\sigma}\right] I(t \geq 0) \quad (10)$$

where  $\Phi$  is the CDF of the standard normal distribution (standard Gaussian),  $\mu$  is the logarithm of the median time of failure time and  $\sigma > 0$  is the shape parameter (deviation parameter). The failure time distribution function based on the Dirichlet Process Lognormal Mixture (DPLNM) is expressed by

$$F(t) = \int \Lambda_{LN}(t|\mu, \sigma^2) G(d\mu, d\sigma^2) \quad (11)$$

where  $G \sim DP(\nu G_0)$  and the base distribution  $G_0$  is considered as prior distribution depending on  $\mu$  and  $\sigma^2$  [24]. Using the Lognormal kernel distribution, the conjugate prior distributions for  $G_0$  distribution parameters,  $\mu$  and  $\sigma^2$ , are assumed Normal and Inverse-Gamma (IG), respectively [26].  $G_0$  for the DPLNM depends on hyper parameters which are incorporated into the hierarchical form of the DPLNM model.

#### B. MCMC Algorithm

Based on the conditional distributions and the prior distributions, we can obtain the full conditional posterior distribution of the Lognormal kernel distribution, which depends on the prior distributions of the hyper parameters. As the prior distribution is a DP the obtained posterior distribution is also a DP as well [22]. The DP mixture model requires sampling through Markov-Chain-Monte-Carlo (MCMC) methods, which can be computationally intensive because updating the Markov Chain process required for each replication is necessary until convergence is achieved. Finally, an estimate  $\hat{G}$  of  $G$ , the empirical distribution density function,

is obtained. Using  $\hat{G}$  and Bayes law, the cumulative distribution function of the failure time is estimated by

$$\hat{F}(t) = \int \Lambda_{LN}(t|\hat{\mu}, \hat{\sigma}^2) \hat{G}(d\mu, d\sigma^2) \quad (12)$$

Then, the PDF of the failure time is

$$\hat{f}(t) = \frac{\partial \hat{F}(t)}{\partial t} = \int \frac{\partial \Lambda_{LN}(t|\hat{\mu}, \hat{\sigma}^2)}{\partial t} \hat{G}(d\mu, d\sigma^2) \quad (13)$$

and consequently, the hazard rate function is obtained by substituting  $\hat{f}(t)$  and  $\hat{F}(t)$  in (2).

The nonparametric Bayesian DPMM using the MCMC procedure of the hierarchical structure (9) is presented in Figure 1.

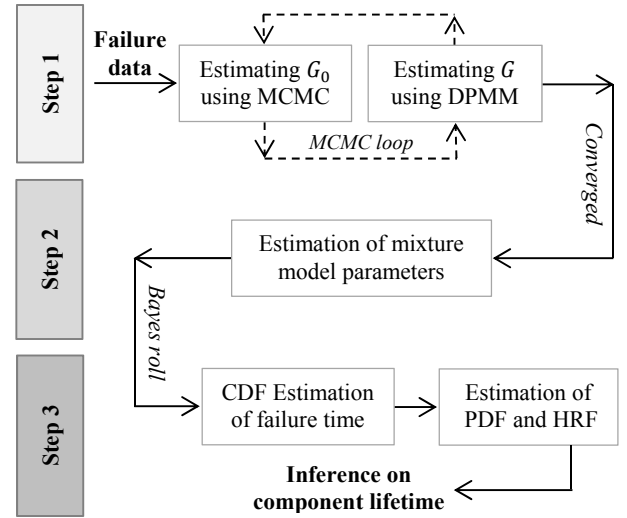


Fig.1 The process of nonparametric Bayesian analysis based on DPMM using MCMC algorithm

### IV. EMPIRICAL RESULTS AND DISCUSSIONS

In order to demonstrate the validity of the proposed models and determine their suitability for railway system reliability analysis, a simulation test is undertaken to study the estimation of the lifetime distributions by the proposed models against a known failure data distribution. The study will further investigate the best kernel to estimate the hazard rate in reliability analysis.

In this test, an artificially generated lifetime data set is provided as reference data in a model-based simulation. Different kernel distributions are examined i.e. Exponential, Weibull and Lognormal to account for different simulation scenarios and their suitability are examined. The data set is generated from complex models, i.e. finite mixture models and competing risk models which have been used in the reliability context [20] and consistent with component lifetime data in railway systems.

The complete data set is of the size  $n = 200$  as the lifetime observations. Non-informative distributions for generating lifetime data sets are selected to avoid further assumptions. These distributions are commonly deployed and they resemble the nature of complex models well in reliability analysis for generating failure time data [26].

Generally, in a mixture model each mixture component density represents the probability distribution function of a group of individuals in the whole population. Based on the nature of the lifetime type data, a mixture of long tail distributions is assumed [17]. A mixture of Lognormal (LN) distribution and Inverse-Gaussian (IGN) distribution is considered. This mixture has a long tail which can be controlled by dispersion parameters of each mixture component. Based on the mixture model definition provided in [13], the  $p$ -LN-ING mixture model is shown as below

$$p \cdot LN(4, 0.16) + (1 - p) \cdot IGN(8, 0.49) \quad (14)$$

where  $p$  is the component's influence ratio and here  $p = 0.3$  is assumed. It means that 30% of the mixture comes from LN and the remaining 70% from ING. This model is also called bi-model [13]. The graphical and numerical results are obtained by the Linux version of R programming language primarily with the "mixtools" and "KernSmooth" packages [27].

The use of simulated data allows examination of the distribution of the data and then the comparison with the obtained results of the parametric finite mixture model (PARMIX) and kernel-based DPMM with it as the reference. Figure 2 illustrates both the actual and estimated PDF and CDF of the generated data based on the mixture model (14) and the comparisons results.

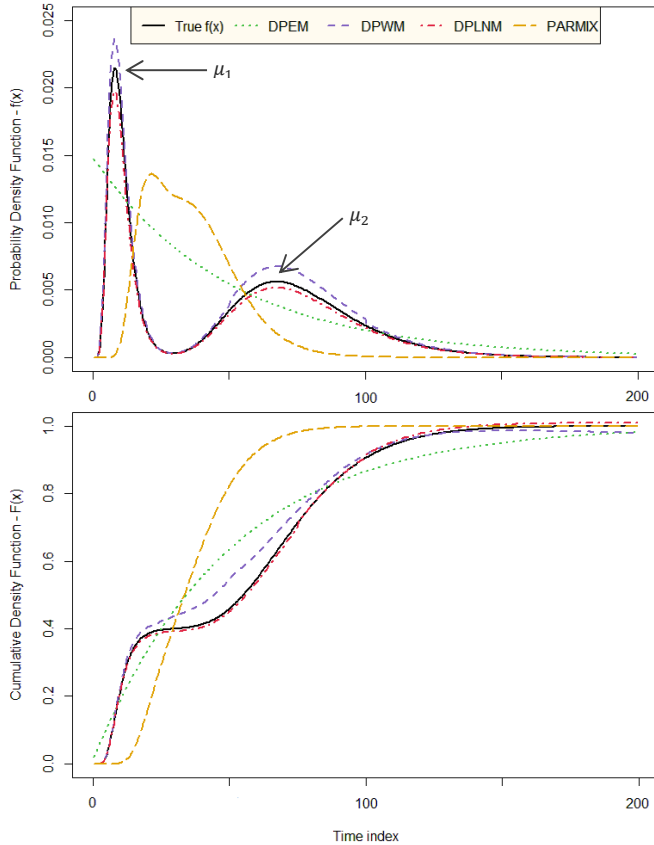


Fig.2 PDF and CDF of the actual data, DPEM, DPWM, DPLNM and PARMIX for simulated lifetime data using mixture model

The performance of PARMIX, Exponential kernel (DPEM), Weibull kernel (DPWM) and Lognormal kernel (DPLNM) and is also compared with those of the actual model. Figure 2 shows that the Exponential kernel and PARMIX are not capable of capturing the generated mixture distribution with long tail. The estimated PDF and CDF of the generated data based on DPWM have the same trend with the actual mixture model with significant differences around the most frequent values of the mixture components ( $\mu_1$  and  $\mu_2$ ), which are indicated with arrows in Figure 2.

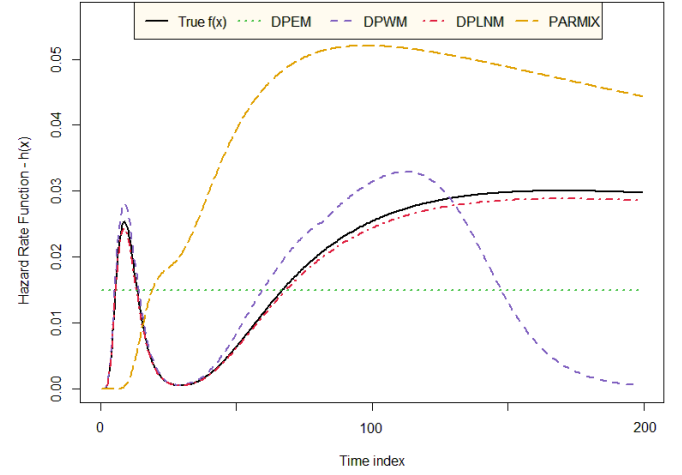


Fig.3 Hazard rate functions of the actual data, DPEM, DPWM, DPLNM and PARMIX for the simulated lifetime data

Figure 3 also illustrates the performance of the parametric method and nonparametric Bayesian method with different mixture models in estimating the hazard rate function. It is clear that PARMIX and DPEM do not match the actual hazard rate function. It is however interesting that the estimated hazard rate function, which is calculated based on estimated PDF and CDF, cannot fit the hazard rate function of the actual mixture model properly. Thus, the DPMM based Weibull kernel (DPWM) is not a good choice for estimating the mixture lifetime data although Weibull kernel is an asymmetric long tail distribution kernel.

The estimated PDF, CDF and hazard rate function of the generated data using DPLNM show a close match. Hence, the Lognormal kernel function, which is a long tail distribution with sharp concentration around the mean, can be a good choice for the lifetime data of mixture type.

The above results are derived from visual inspection of graphs. To facilitate a quantitative comparison, the Kolmogorov-Smirnov test, a nonparametric test for goodness-of-fit [28], is employed here to investigate the goodness-of-fit of the mixture models against the assumed mixture model. The null hypothesis of the Kolmogorov-Smirnov test states that the estimated cumulative density (or a function of the cumulative density) equals the assumed density imposed by the proposed mixture model. The alternative hypothesis is that the true density is not equal to the mixture model density. Table 1 depicts the Kolmogorov-Smirnov test statistics as well as the p-values of the studied mixture models.

The results in Table I show that the estimated CDF for the mixture model using DPLNM has the smallest test statistics value of 0.02681 with a p-value of  $0.868 >> 0.05$ . It can be concluded the hypothesized cumulative density, as well as the probability density function (PDF) based on DPLNM, is a good estimate. This is supported by the significant p-value of  $0.868 >> 0.05$ . For the density estimation, DPWM also performs reasonably well (p-value of 0.106), but the hazard rate can only be well approximated by DPLNM, which is the only method to give a non-significant p-value of  $0.667 > 0.05$ .

TABLE I

Kolmogorov-Smirnov goodness-of-fit test of failure time cumulative density and hazard rate function estimation

Models	Kolmogorov-Smirnov test			
	$F(x)$		$h(x)$	
	Test Stats	P-value <sup>a</sup>	Test Stats	P-value <sup>a</sup>
PARMIX	0.1652	0.004	0.2193	0.000
DPEM	0.1946	0.002	0.2164	0.000
DPWM	0.1106	0.106	0.0694	0.039
DPLNM	0.02681	0.868	0.0451	0.667

<sup>a</sup> The significant level is 0.05

## V. CONCLUSIONS

This study has investigated the application of a non-parametric Bayesian mixture approach with the Dirichlet Process Mixture Model (DPMM) through a number of the kernel densities in the context of reliability analysis for a railway system at component level. The results illustrated that this method offers significant flexibility to account for complicated mixture distributions for which parametric methods often fail. The results also show that the Log-normal kernel is preferred over the Gaussian and Weibull kernels to lead to a reasonably good estimate of the hazard rate of the components, which is one of the primary requirements of reliability analysis.

The results here pave the way for reliability analysis at a higher level on which the component functionalities and dependencies contribute collectively toward the specified safety and availability requirements of the integrated system. It then facilitates further studies on maintenance planning, failure consequences and life cycle cost for a railway system.

## REFERENCES

- [1]. L.C. Wolstenholme, *Reliability modelling: a statistical approach*. Chapman and Hall CRC, 1999.
- [2]. C.R. Avery, "Power electronics reliability in rail traction", *Proc. IEEE Colloquium Power Electronics Reliability – Promise and Practice*, London, UK, pp. 1-7, 1998.
- [3]. S.K. Chen, T.K., Ho and B.H. Mao, "Reliability evaluations of railway power supplies by fault-tree analysis", *IET Electr. Power Appl.*, vol. 1(2), pp. 161-172, 2007.
- [4]. F. De Felice and A. Petrillo, "Methodological approach for performing human reliability and error analysis in railway transportation system", *International Journal of Engineering and Technology*, vol. 3(5), pp. 341-353, 2011.
- [5]. M. Akama, "Bayesian analysis for the results of fatigue test using full-scale models to obtain the accurate failure probabilities of the Shinkansen vehicle axle", *Reliability Engineering and System Safety*, vol. 75, pp. 321-332, 2002.
- [6]. A.R. Andrade and P.F. Teixeira, "A Bayesian model to assess rail track geometry degradation through its life-cycle", *Research in Transportation Economics*, vol. 36, pp. 1-8, 2012.
- [7]. E.R. Vaidogas, "Bayesian approach to forecasting damage to building from accidental explosions on railway", *Transport and Telecommunication*, vol. 7(1), pp. 34-43, 2006.
- [8]. C. Wang, X. Liao, L. Carin and D.B. Dunson, "Classification with incomplete data using Dirichlet Process Priors", *Journal of Machine Learning Research*, vol. 11, pp. 3269-3311, 2010.
- [9]. T.E. Hanson, "Modeling censored lifetime data using a mixture of gammas baseline", *Bayesian Analysis*, vol. 1(3) pp. 575-594, 2006.
- [10]. E. Gelfand and A. Kottas, "A computational approach for full nonparametric Bayesian inference under Dirichlet Process Mixture Models", *Journal of Computational and Graphical Statistics*, vol. 11(2), pp. 289-305, 2002.
- [11]. A. Kottas, "Nonparametric Bayesian survival analysis using mixtures of Weibull distributions", *Journal of Statistical Planning and Inference*, vol. 136(3), pp. 578-596, 2006.
- [12]. J. I. Ansell and M. J. Phillips, *Practical Methods for Reliability Data Analysis*. Oxford Science Publications, New York, 1994.
- [13]. G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley, New York, 2000.
- [14]. J. Diebolt and C.P. Robert, Estimation of Finite Mixture Distributions through Bayesian Sampling, *J.R. Statist. Soc. B*, vol. 56(2), pp. 363-375, 1994.
- [15]. M. Hollander and D.A. Wolfe, *Nonparametric Statistical Methods*, 2nd Edition, Wiley-Interscience, 1999.
- [16]. E. Antoniak, "Mixtures of Dirichlet Process with applications to Bayesian nonparametric problems", *The Annals of Statistics*, vol. 2(6), pp. 1152-1174, 1974.
- [17]. N. D. Singpurwalla, *Reliability and risk: a Bayesian perspective*. England: John Wiley & Sons, 2006.
- [18]. R. L. Dykstra and P. Laud, "A Bayesian nonparametric approach to reliability", *The Annals of Stats*, vol. 9(2), pp. 356-367, 1981.
- [19]. Y. Lo, "On a class of Bayesian nonparametric estimates: I. Density estimates", *The Annals of Stats*, vol. 12(1), pp. 351-357, 1984.
- [20]. Kottas, "Nonparametric Bayesian survival analysis using mixtures of Weibull distributions", *Journal of Statistical Planning and Inference*, vol. 136(3), pp. 578-596, 2006.
- [21]. D. Belyi, P. Damien, E. Popova and D. Morton, "Bayesian nonparametric simulation of hazard functions", in *Proc. Winter Simulation Conference (WSC)*, pp. 475-482, 2009.
- [22]. T. S. Ferguson, "A Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, vol. 1(2), pp. 209-230, 1973.
- [23]. J. D. McAuliffe, D. M. Blei, M. I. Jordan, "Nonparametric empirical Bayes for the Dirichlet Process mixture model", *Statistical and Computing*, vol. 16, pp. 5-14, 2006.
- [24]. M. D. Escobar, "Estimating normal means with a Dirichlet Process prior", *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 268-277, 1994.
- [25]. P. Müller and F. A. Quintana, "Nonparametric Bayesian data analysis", *Statistical Science*, vol. 19(1), pp. 95-110, 2004.
- [26]. M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures", *Journal of the American Statistical Association*, vol. 90, pp. 577-588, 1995.
- [27]. The comprehensive R archive Network. [www.cran.r-project.org](http://www.cran.r-project.org).
- [28]. W. Silverman, *Density estimations for statistics and data analysis*. Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.